Scale Counting

S.H. Steiner & R.J. Mackay

IIQP Research Report
RR-01-07

Scale Counting

Stefan H. Steiner and R. Jock MacKay
Dept. of Statistics and Actuarial Sciences
University of Waterloo
Waterloo, N2L 3G1 Canada

In many applications small parts, such as nuts and bolts, are counted using a scale. With scale counting the number of parts is estimated by weighing them and dividing the total weight by the (estimated) average weight of an individual part. This procedure avoids counting individual parts and can thus save time and money and improve the accuracy of counts. In this article the effect of the estimation procedure used to determine the average weight, sample size, measurement error, and measurement resolution on the accuracy of the scale count are explored and quantified. General rules of thumb that suggest when scale counting is likely to be beneficial are presented. Changes in the standard implementation of scale counting are suggested.

1. Introduction

In many applications small parts are counted using a weighing method, called scale counting. See Figure 1 for an illustration, and Anonymous (1981) for more background. Using scale counting we count the parts by weighing them and dividing the total weight by the (estimated) average weight of each individual part. This procedure avoids counting individual parts and can thus save time and money and improve the accuracy of counts. However, as far as we are aware, the statistical properties of scale counting have never been extensively studied.

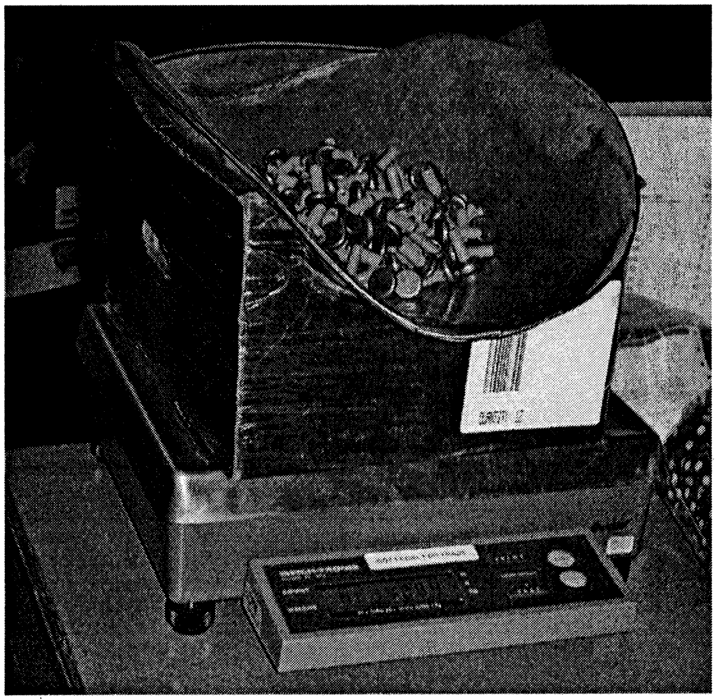

Figure 1: Counting Scale in Action

The effectiveness of scale counting, measured in terms of the accuracy of the count, depends on the variability of the weights of the individual parts, the measurement bias, variability and resolution, the number of parts being counted, and the procedure used to estimate the average weight of individual parts.

This article was motivated by an application in the automotive industry. In the application the goal is to package automotive parts in crates for shipment overseas (where the vehicles are assembled). The warehouse from which the shipments originate contains over 3000 different parts or components. The parts range in cost and size from engines to small fasteners. The parts are shipped in crates that are designed to contain all parts on a particular list of parts needed to build around 100 vehicles. In this application obtaining the correct number of parts of each type in each crate is extremely important. If there are too few of any component not all the vehicles can be assembled, whereas too many of particular component leads to waste, or worse if attempts are made to use the excess components in some other assembly operation. The parts needed for a particular crate are loaded by pickers who roam the warehouse adding all the parts on their pick list. The number of pieces needed of each part varies from around 100 pieces to over 3500 pieces (e.g. some common bolts). Currently many of the parts are hand counted, but some of the smaller and cheaper parts are counted using scales. Specific aspects of this example will be explored in more detail in this article.

This article is organized in the following manner. First, in Section 2, we detail the current standard scale counting procedures. We explain the calibration procedures and discuss the different goals of scale counting achieved through bulk and dribble counting. We also develop some approximations that show how the accuracy of the scale counting results depends on the number of pieces used in the calibration step, the number of parts to be counted, and the coefficient of variation for the individual part weights. The results in Section 2 are derived assuming a perfect weighing device and ignoring the discretization needed to yield an integer count. In Section 3, we quantify the effect of relaxing these assumptions, and provide guidelines when measurement error and/or the discretization can substantially effect the count. Section 4 we turn to issues of importance to implementation of scale counting. We make some recommendations for changes to the standard practice. In addition, a new procedure, called multiple scale counting is proposed that yields much more accurate counts than the standard procedure in some circumstances. Finally, in Section 5 we summarize our results.

2. Current Scale Counting Approaches

Consider two different goals when using scale counting. First, we may be interested in determining the number of parts in a given group. To estimate the number of parts using a weighing scale all the parts are placed on the scale at once, and the total weight is divided by the (estimated) average weight of the parts to obtain an estimate for the total number of parts. This is called bulk counting. Second, we may wish to create a group (or groups) of a specified number of parts. This second goal can be accomplished by adding and removing parts from the scale until the estimated total number of parts (again derived by dividing the total weight by the estimated average weight) equals the desired value. This second goal has the descriptive title of dribble counting. For both scale counting goals, an estimate for the average weight of the parts is required. The average part weight is estimated using a calibration step where a small number of parts are *manually counted* and weighed. We discuss the calibration step of scale counting in Section 2.1. Although the two procedures, bulk and dribble counting have different goals, we shall show that the analysis to determine how well these goals can be attained is the same.

To set notation and to simplify the analysis we assume the individual parts have true weights that follow a normal distribution. This is denoted $X_i \sim N(\mu, \sigma^2)$, where μ and σ are the mean and standard deviation of the individual part weights respectively. In addition, we assume the weights of the individual parts are independent.

2.1 Calibration Step

The precision of the estimate for the average weight is crucial to the success of scale counting. It is possible to use an average weight derived from historical records, but usually, like in our motivating example, due to concerns about a possibly drifting mean weight, a small sample of the current parts is used.

In the calibration step we count out p parts and determine their total weight. We estimate the average part weight, denoted $\hat{\mu}$, by dividing total weight of those p parts by p, i.e. $\hat{\mu} = \hat{t}_p / p$, where $\hat{t}_p$ is the estimated total weight of the p parts. Usually the calibration sample size p will be much smaller than the number of parts we ultimately wish to count. Then, for the moment ignoring measurement error, the corresponding estimator, denoted $\tilde{\mu}$, has distribution $N(\mu, \sigma^2/p)$. For more information on the effect of measurement error see Section 3.2. Note that by using this calibration procedure it is not possible to estimate the variability in the individual weights. This is of concern since, as will be shown later, the variability in the individual weights, through the coefficient of variation σ/μ, is a crucial parameter that strongly influences the effectiveness of scale counting.

In the motivating automotive example a calibration sample size of 25 parts (i.e. $p = 25$) is used for all scale counting. Due to possible variability from shipment to shipment in μ this calibration is repeated whenever a picker needs to count some parts. For simplicity, the calibration is repeated even if the parts come from a shipment batch that has been used earlier. Note that p equals 25 is chosen regardless of the coefficient of variation σ/μ of the individual part weights. This is done for simplicity, but also because the large number of different parts makes estimating all the different coefficients of variation a daunting task.

2.2 Bulk Counting

Say we wish to count a group of roughly n parts, where n is unknown. Assume we have previously estimated the average part weight using a sample of p parts as described in Section 2.1. Then, an estimate of the total number of parts is

$$\hat{n} = \hat{t}_n / \hat{\mu}, \tag{1}$$

where $\hat{t}_n$ is the measured (by the scale) total weight of all n parts and $\hat{\mu}$ is the estimated average part weight given in Section 2.1. In practice the results from (1) are rounded off to the nearest integer. For the moment we ignore the effect of rounding. In Section 3.1 we explore the effect of rounding in more detail, and show that, except when the variability in the count is very small the effect of rounding is not important. Assuming no measurement error, the estimator corresponding to the estimate $\hat{t}_n$, denoted $\tilde{t}_n$, has distribution $N(n\mu, n\sigma^2)$. Thus, the estimator corresponding to $\hat{t}_n/\hat{\mu}$, denoted $\tilde{t}_n/\tilde{\mu}$, will have distribution $N(n\mu, n\sigma^2)/N(\mu, \sigma^2/p)$. There is no dependency between $\tilde{t}_n$ and $\tilde{\mu}$ since we assume they come from two different samples and we assume independence between the individual weights.

The probability density function of the estimator $\tilde{t}_n/\tilde{\mu}$ is the ratio of two normally distributed variables. We may approximate the distribution of $\tilde{t}_n/\tilde{\mu}$ using method proposed by Cabuk and Springer (1990). The estimator is not unbiased, since its expected value is not n, and it is not normally distributed. However, in most cases of interest in this application, i.e. where n and p are large (say larger than 100 and 5 respectively) and $\sigma/\mu = \gamma$, the coefficient of variation, is small (say smaller than .05) the estimator is approximately unbiased, has only negligible skewness, and is closely approximated by a normal distribution. If we assume approximate normality we need only estimate the mean and standard deviation of $\tilde{t}_n/\tilde{\mu}$. This can be accomplished through the method of statistical differentials (i.e. Taylor series expansion) retaining terms up to second order (Kotz and Johnson, 1982). We obtain

$$E(\tilde{t}_n/\tilde{\mu}) \cong \frac{E(\tilde{t}_n)}{E(\tilde{\mu})}\left[1+\frac{Var(\tilde{\mu})}{E(\tilde{\mu})^2}\right] = n\left(1+\frac{\gamma^2}{p}\right) \quad \text{and} \tag{2}$$

$$Var(\tilde{t}_n/\tilde{\mu}) \cong \left[\frac{E(\tilde{t}_n)}{E(\tilde{\mu})}\right]^2\left[\frac{Var(\tilde{t}_n)}{E(\tilde{t}_n)^2}+\frac{Var(\tilde{\mu})}{E(\tilde{\mu})^2}\right] = n\gamma^2\left(1+\frac{n}{p}\right) = \sigma_r^2 \tag{3}$$

where $\delta_r = n\gamma^2/p$ and σ_r are the bias and standard deviation of the continuous estimator. Note that the bias is always positive. The approximations (2) and (3) have been found to quite accurate for cases of interest in this application. In some implementations of bulk counting, we add to the calibration sample rather than starting a fresh. The bias and standard deviation in the count that arise in this case are equivalent to what we would get if we simply bulk count *n-p* parts.

It is apparent from (2) and (3) that the effectiveness of scale counting depends strongly on the variability in the individual part weights (quantified by the coefficient of variation of the individual part weights γ) and the calibration sample size, *p*. This dependency is shown pictorially in Figures 2 and 3. Figure 2 plots contours of the percent bias, given in terms of a percentage of the total number of parts (i.e. $100\gamma^2/p$). We see that the percent bias is very small unless σ/μ is large and *p* is very small. The bias is at most 0.05% if we assume $\gamma < 0.05$ and $p > 5$.

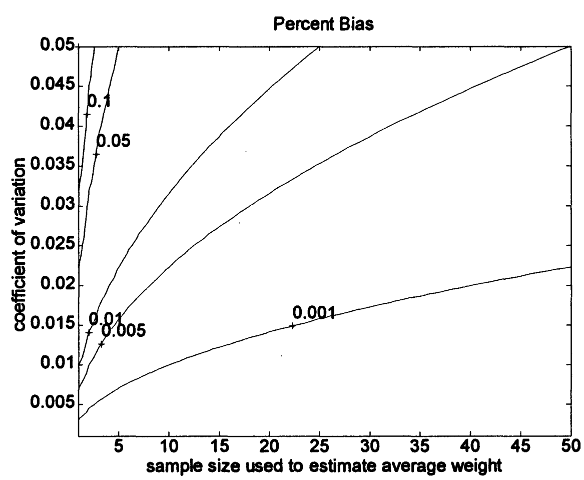

Figure 2: Percent Bias of the Estimated Total Number of Parts

The variance of the estimator for the total number of parts, also expressed as a percentage of *n*, is a function of just the coefficient of variation and the percent of the total sample used in the calibration step, i.e. $100p/n$. In most applications $p \ll n$. Figure 3 shows contours of the percent variance $100\gamma^2(1+n/p)$. The variance can be large, especially if a small calibration sample size is used.

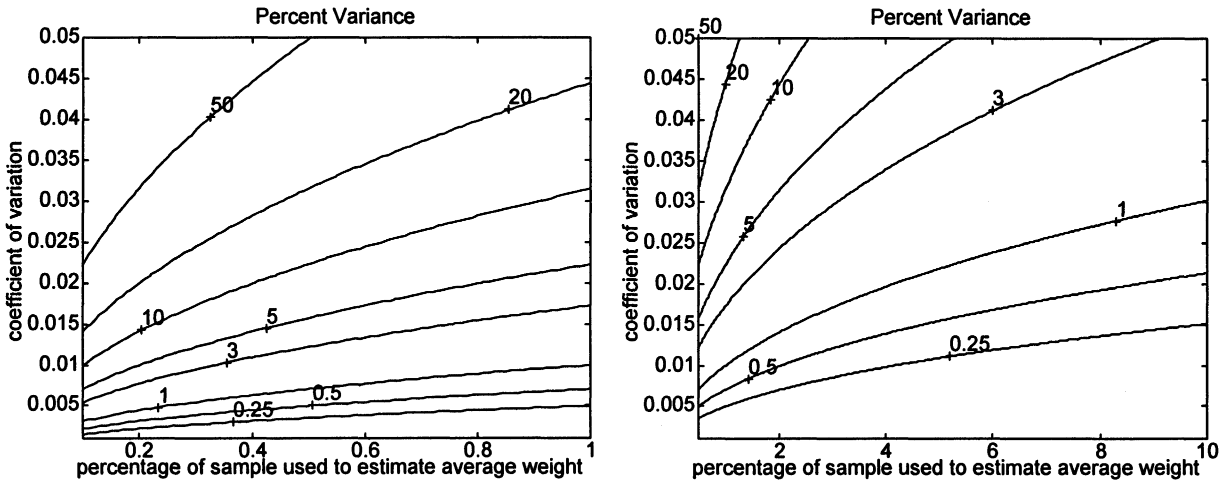

Figure 3: Percent Variance of the Estimated Total Number of Parts

Consider an example where the actual number of parts is close to n=800, p=25 (i.e. around 3.1% of sample is used to estimate the average part weight), μ=.15 and σ=.005 (i.e. coefficient of variation is γ=.033). Then, from (2) and (3) the expected bias and standard deviation in the scale count are approximately .036 and 5.4 units respectively. In this example, the bias is exceedingly small, but the standard deviation in the scale count is very large, and we will not consistently be able to accurately count the number of parts.

2.3 Dribble Counting

In dribble counting, the goal is to create a group or groups of n parts rather than to count the number of parts in a group. This goal is achieved by adding and possibly removing parts from the scale until the correct count is obtained. An automatic scale counting machine we have seen used in practice dumps parts onto the scale until it contains close to the correct number of parts. Then, additional parts are added one at a time until the scale first concludes there are n (or greater) parts. More general procedures that allow removing parts are also possible. However, the dribble procedure makes little difference to the statistical properties of the results.

With dribble counting the random variable is the actual number of parts on the scale, while with bulk counting the random variable is the scale count. We can write the probabilities of different outcomes in dribble and bulk counting as follows. Assume n is the target number of items for the dribble counting procedure. Let y be the actual number of parts on the scale at the end of the dribble counting, and let Y be the corresponding random variable. Then the distribution of Y is discrete and $\Pr(Y = y) = \Pr(sc(y) = n)$, where $sc(y)$ is the estimated number of parts we obtain from bulk counting y parts. Similarly, when bulk counting n parts the distribution of the scale count is $\Pr(Y = y) = \Pr(sc(n) = y)$. When y is small, dribble counting and bulk counting are statistically equivalent because both procedures will always yield the correct result. For large y, dribble counting and bulk counting are approximately equivalent in terms of their statistical properties because, assuming σ is small enough that scale counting a single item always yields unity, through symmetry $\Pr(sc(y) = n) \cong \Pr(sc(n) = y)$. Thus, all the results from Section 2.2 derived for bulk counting are equally applicable to dribble counting.

3. Effects of Measurement Error, Resolution and the Discretization

In Section 2 we derived approximations for the bias and variability of the scale count results under the assumption of a perfect measurement device, and ignoring the rounding off of the count to an integer. In this section, we explore the effect of relaxing those assumptions on the accuracy of the scale count. First, in Section 3.1 the effect of rounding off to integers is examined more closely. It is shown that unless the variability in the scale count is very small the effect of the discretization can be ignored or modeled using a simple approximation. In Sections

3.2 and 3.3 we quantify the effects of measurement error and measurement resolution respectively. Using these results it is possible determine when the properties of the measurement device may negatively effect scale counting.

3.1 Effect of Discretization

The scale count is always an integer. That is, the estimated number of items is $round\left(\hat{t}_n/\hat{\mu}\right)$. Here we examine the effect of the discretization on the properties of the estimator $\hat{t}_n/\hat{\mu}$ studied in Section 2. The effect of this discretization is not always intuitive. In some cases the rounding off greatly reduces the variability, while in most cases the variability increases slightly.

As suggested in Section 2.2, $\tilde{t}_n/\tilde{\mu}$ is approximately normally distributed with bias and standard deviation δ_r and σ_r respectively i.e. $\tilde{t}_n/\tilde{\mu} \sim N\left(n+\delta_r, \sigma_r^2\right)$. For the discretization, the magnitude of n makes no difference since we are rounding to the nearest integer. As such, to explore the effect of the discretization generally we consider $round(X)$, where $X \sim N\left(\theta_r, \sigma_r^2\right)$, $\theta_r = \delta_r - round\left(\delta_r\right)$, the non-integer part of the bias, and $-.5 \le \theta_r \le .5$. We denote the mean and standard deviation of $round(X)$ as δ_d and σ_d respectively. The resulting bias in any given problem is then given by $round\left(\delta_r\right) + \delta_d$. In many applications $round\left(\delta_r\right)$ is zero. To determine the effect of the discretization we compare δ_d and σ_d with θ_r and σ_r. We may calculate δ_d and σ_d easily from the resulting discretized normal distribution. For example, assume δ_r equals zero, then we get $\sigma_d = 0.03$ when $\sigma_r = 0.15$, and $\sigma_d = 0.57$ when $\sigma_r = 0.5$. These two examples illustrate that the effect of the discretization can either increase or decrease the standard deviation. Figure 4 shows contours of δ_d and σ_d when $-.5 \le \theta_r \le .5$ and $0 < \sigma_r \le 1$.

From Figure 4 we see that for large σ_r the discretization has no effect on the bias, i.e. δ_d = θ_r, however, when both θ_r and σ_r are small the discretization greatly reduces the bias, i.e. δ_d << θ_r. Similarly, for large σ_r the discretization results in a slight increase in the standard deviation but, when both θ_r and σ_r are small the discretization greatly reduces the standard deviation, i.e. $\sigma_d << \sigma_r$.

We may closely approximate the effect of the discretization in most case by modeling $round(\tilde{t}_n/\tilde{\mu})$ as $\tilde{t}_n/\tilde{\mu}+U$, where U is a continuous uniform random variable that ranges between $-1/2$ and $1/2$. It is well known that $E(U) = 0$ and $Var(U) = 1/12$. This approximation for the rounding is good when the variance of $\tilde{t}_n/\tilde{\mu}$ is large, say greater than 0.5. Using this approximation $\sigma_d \cong \sqrt{\sigma_r^2+1/12}$.

Bias of Discretized Distribution   Standard Deviation of Discretized Distribution

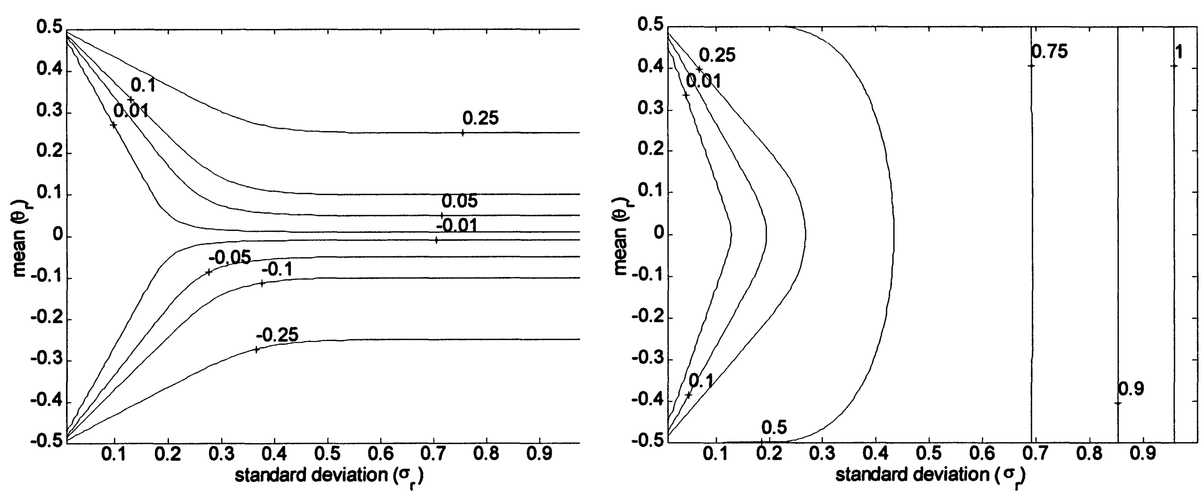

Figure 4: Plots of the Mean and Standard Deviation of *round*(X) as a Function of δ_r and σ_r

Note that from Figure 2 the bias δ_r is typically very small. Also, from Figure 3, in most cases of practical interest the magnitude of σ_r is large relative to the units of measurement and thus the discretization in the scale counting procedure has little effect. However, there are potentially dramatic beneficial effects of discretization when both δ_r and σ_r are small. This property will be exploited in Section 4.3 where we introduce the idea of multiple scale counting.

3.2 Effect of Measurement Error

To derive the expressions (2) and (3) we assumed no measurement error. In most applications this assumption is not realistic. We now relax the assumption of no measurement error, and show that the negative effect of measurement error is usually very small, or can be avoided.

Suppose the measurement bias and standard deviation are δ_m and σ_m respectively. Then, assuming normally distributed measurement errors we have $\tilde{\mu} \sim N\left(\mu + \delta_m/p, \left(p\sigma^2 + \sigma_m^2\right)/p^2\right)$ and $\tilde{t}_n \sim N\left(n\mu + \delta_m, n\sigma^2 + \sigma_m^2\right)$, with $\tilde{t}_n$ and $\tilde{\mu}$ independent. Applying the general results from statistical differentials given by (2) we obtain the approximations

$$E\left(\tilde{t}_n/\tilde{\mu}\right) \cong \frac{n\mu + \delta_m}{\mu + \delta_m/p}\left(1 + \frac{\left[p\sigma^2 + \sigma_m^2\right]}{\left(p\mu + \delta_m\right)^2}\right) \text{ and} \tag{4}$$

$$Var\left(\tilde{t}_n/\tilde{\mu}\right) \cong \left(\frac{n\mu + \delta_m}{\mu + \delta_m/p}\right)^2\left(\frac{n\sigma^2 + \sigma_m^2}{\left(n\mu + \delta_m\right)^2} + \frac{p\sigma^2 + \sigma_m^2}{\left(p\mu + \delta_m\right)^2}\right)$$

We see from (4) that the additional bias and variability in the estimated total number of parts introduced by the measurement variability (σ_m) is small if σ_m is substantially less than σ, and thus that $\sigma_m^2 << p\sigma^2$. This assumption implies that the measurement device does not add much variability over and above the variability in the parts themselves when weighing the calibration sample. However, the effect of measurement bias can be substantial, mostly because δ_m/p may be non-negligible relative to μ. Substantial measurement bias (i.e. δ_m is large) will have a large negative effect on the results mostly by introducing a bias in the total count.

However, we can avoid the problem by eliminating measurement bias through weighing both a container and the parts in a container, and calculating the weight of the parts as the difference between the two observed weights. This procedure is used in the motivating example as illustrated in Figure 1. Using differencing will eliminate the measurement bias, at the expense of a small amount of added variability, so long as the amount of bias does not depend on the weight of the part weighed.

3.3 Effect of Measurement Resolution

All the previous analysis has ignored the possible effect of poor measurement resolution. Measurement resolution is defined by the small unit of measurement. For example, we may measure weights to the nearest two grams. Then, any collection of units that actually weighed between 99 and 101 grams would yield a weight of 100 grams. In situations where the measurement unit is small compared with the expected uncertainty due to measurement variability the measurement resolution has little effect. However, when the measurement unit is greater than say σ_m resolution can have a substantial negative effect on scale counting. We quantify the resolution of a measurement device, denoted *r,* using one over the scale's minimum discrimination weight (or measurement unit). Then, ignoring measurement error, a measurement device with resolution r that weighs a part of weight w yields the result $round(rw)/r$. For example, if a scale has a capacity of 20 kilograms and is capable of 10000 divisions, the scale's minimum discrimination is 2 grams and r is 1/2 (when parts are weighed in grams). Similarly, for example, r =100 means that the measurement device is capable of providing around 2 decimal points.

To determine the effect of measurement resolution on the results of scale count we need to know how the resolution effects the distribution of $\tilde{t}_n / \tilde{\mu}$. We can approximate the effect of the rounding due to the measurement resolution by modeling *round*(X) as $X+U$, where U is a continuous uniform random variable that ranges between $-1/2$ and $1/2$. This is the same approximation used in Section 3.1. Using our definition of *r,* and assuming no measurement error we can write

$$\tilde{t}_n / \tilde{\mu} \cong \frac{round\left[N\left(rn\mu, nr^2\sigma^2\right)\right]/r}{round\left[N\left(rp\mu, pr^2\sigma^2\right)\right]/pr} \cong \frac{p\left[N\left(rn\mu, nr^2\sigma^2\right)+U\right]}{N\left(rp\mu, pr^2\sigma^2\right)+U}. \tag{5}$$

Applying the method of statistical differentials, as in (2) and (3), to the ratio (5) we get

$$E\left(\tilde{t}_n/\tilde{\mu}\right) \cong n\left(1+\frac{\sigma^2}{p\mu^2}+\frac{1}{12p^2r^2\mu^2}\right) \text{ and} \tag{6}$$

$$Var\left(\tilde{t}_n/\tilde{\mu}\right) \cong n\left(\frac{\sigma^2}{\mu^2}+\frac{1}{12nr^2\mu^2}+\frac{n\sigma^2}{p\mu^2}+\frac{n}{12p^2r^2\mu^2}\right)$$

In any application we can quantify the *expected* negative effect of poor resolution on the bias and variability of the count by comparing (6) with (2) and (3). However, in any application the effect of poor measurement resolution can be larger than the expected difference if $r\mu p$ happens to lie halfway between two integers. This is especially important if $r\mu p$ is small. It is unfortunately not possible to predict how often this worst case would be achieved since we have no control over μ.

The differences between the expressions given by (6) and those given by (2) and (3), where we ignored the effect of measurement resolution, are the terms that include the parameter *r.* Since generally $n >> p$, we know $\frac{n}{12p^2r^2\mu^2} >> \frac{1}{12nr^2\mu^2}$. Also, typically $\frac{n\sigma^2}{p\mu^2} > \frac{\sigma^2}{\mu^2}$. Thus, to bound the effect of the measurement resolution on both the bias and variability of the count we need to compare $\frac{n\sigma^2}{p\mu^2}$ and $\frac{n}{12p^2r^2\mu^2}$. Limiting the variance and bias in the scale count introduced by poor resolution to be smaller (say less than half the size) than that introduced by errors in the estimation of the average part weight seems a reasonable goal. With that goal in mind we want the measurement resolution to be good enough so that $(r\mu)^2 > 2\frac{1}{12p} \Big/ \left(\frac{\sigma}{\mu}\right)^2$ or equivalently $r > \frac{2}{\sigma\sqrt{12p}}$.

In the motivating example, the minimum discrimination of the scales used is two grams. Thus, when measuring parts in grams the measurement resolution r is 0.5. The calibration step uses p equal to 25 parts. Assuming a coefficient of variation equal to 0.025 the rule of thumb introduced in the previous paragraph implies we want $r\mu > 4.6$, i.e. the average weight of individual parts should be greater than around 9 grams. Otherwise the effect of measurement resolution on the scale count results is substantial.

4. Implementation of Scale Counting

This section addresses some issues in the practical implementation of scale counting. First, as discussed in Section 2, the coefficient of variation of the individual part weights is critical in understanding the effectiveness of scale counting, but is not estimated using the standard procedures. In Section 4.1 we discuss alternative strategies for estimating the coefficient of variation. Second, in Section 4.2, we explore how mistakes in the hand count of the calibration sample can influence the scale count. In addition, we suggest a simple checking procedure that can quickly verify the hand count. Third, in Section 4.3, we consider how different consequences of over or under estimating the actual number of parts when scale counting can be accommodated by adjusting the target. Finally, Section 4.4 proposes a new scale counting procedure that uses multiple scale counts. Multiple scale counting is shown, in some circumstances, to greatly reduce the bias and variability in the final count while slightly increasing the complexity of the procedure. Guidelines are provided showing when multiple scale counting should be considered.

4.1 Estimating the Variability of Individual Part Weights

As shown in Section 2 the bias and variability of the scale count are highly dependent on the coefficient of variation of the individual part weights $\gamma = \sigma / \mu$. With the standard calibration procedure, as described in Section 2.1, a single group of parts is weighed. However, to estimate the variability of the individual weights we must weight a number of parts (or groups of parts) separately. One possible approach to obtaining an estimation of γ is to change the calibration step so that each of the p parts is weighed individually. This small change in procedure would require somewhat more effort, but would not effect the precision of the estimate for the average part weight if there is no measurement error or resolution problems. However, this may not be a good idea since when weighing individual parts the effect of poor measurement resolution and measurement error will be more pronounced than when weighing a group of p parts. Also, there is probably no need to re-estimate γ as often μ since the coefficient of variation is not needed to determine a scale counting result. The value of γ relevant more from a planning perspective when making decisions such as how large a calibration sample is necessary to ensure the desired count accuracy. Knowing γ would allow derivation of a confidence interval for the scale count result. Generally what is needed is some idea of the magnitude of γ for different parts and lots of the same parts. As such, we recommend a separate study to estimate γ. For this separate study we would weigh a number of parts individually, and estimate the mean and standard deviation (and thus γ) using the sample mean and standard deviation of the observed weights. If the results in Sections 3.2 and 3.3 suggest there will be a substantial negative effect due to measurement problems when weighing individual parts a possibility is to estimate the variability in individual part weights by weighting many groups of parts. The size of the groups could be chosen so as to reduce the measurement effects to manageable levels.

4.2 Checking the Calibration Sample Size

The results derived in Sections 2 and 3 assume no error is made in the hand-counting of the calibration sample size. An error in the hand-counting can lead to a substantial bias in the scale-counting procedure. For example, assume the calibration sample was suppose to be 25 parts but is actually only 24 parts, then using (2) the bias in the count is around 4%. A larger calibration sample reduces the uncertainty of the estimate for the average part weight but also increases the likelihood of miscounting. Thus, there is an inherent tradeoff. Often in practice procedures are employed that allow a check of the hand-count of the number parts in the calibration sample.

One sample size checking procedure is to add a specified number of additional parts to the scale and verify that the scale count matches the expected total. For instance, in our example application the calibration sample size is intended to be $p = 25$. To check the count the operator adds an another 15 parts and verifies that the scale count reads 40. To assess this hand count checking procedure in the general case we denote the number of additional parts q, and the actual number of parts used in the calibration step as p^*. Using the approximations provided by statistical differentials and some algebra it is possible to show that the distribution of calculated count is approximately normally distributed with mean $p+\frac{pq}{p^*}\left(1+\frac{\gamma^2}{p^*}\right)$ and variance $\frac{qp^2\gamma^2}{\left(p^*\right)^2}\left(1+\frac{q}{p^*}\right)$, where γ is the coefficient of variation for the weights of the individual parts.

We want to choose q large enough so that if p^* does not equal the desired amount, i.e. $p^* \neq p$, the resulting scale count will likely not yield $p+q$. Figure 5 shows contours of minimum value of q needed so that 95% of the time the checking procedure would identify a calibration sample whose size differs from the desired number p.

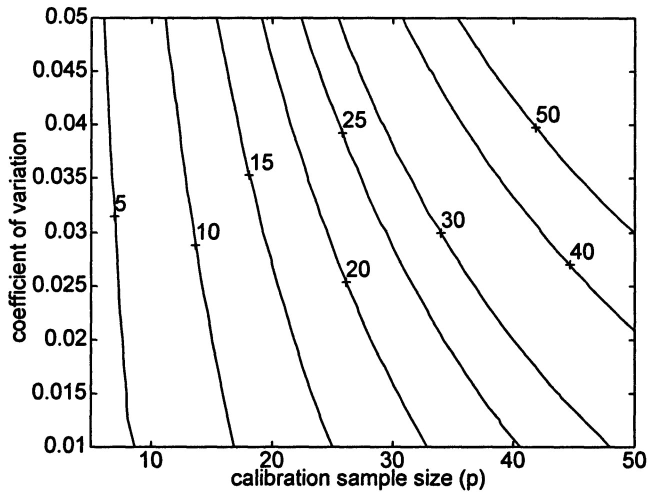

Figure 5: Minimum Additional Size Sample Needed to Check Hand Count

4.3 Asymmetric Loss Functions

Scale counting will not always yield the correct count. The estimated number of parts may be less or greater than the actual number of parts on the scale. The actual number could be either over or under estimated. In some applications of scale counting the consequences of making a mistake may depend on whether the actual number of parts was overestimated or underestimated. This seems especially relevant for dribble counting. For example, in our motivating example packages of 100 or more scale count parts are sent to customers. In this application sending too few parts is a more severe problem than sending too many parts. Similarly, in bulk counting incorrect estimates of the number of parts in a group may result in errors in the inventory count with overestimates and underestimates perhaps not being equally undesirable.

Unequal consequences for over and under estimation suggest an asymmetric loss function. As a result, when dribble counting it often makes sense to aim for more parts than are actually needed to make sure at least the desired number of parts are obtained. Similarly in bulk count we may wish to purposely under report the observed scale count. This idea is employed in our motivating example where the costs associated with picking too few parts is greater than the cost of too many parts. As a result, in the example, specification limits on the actual number of parts picked are set at *n* and *n* plus 4%, and we aim to count *n* plus 2%. In other words, if 500 parts are needed the dribble counting target is set at 510 parts.

The amount by which the target should exceed the actually desired number of parts depends on the variability and bias in the dribble count, and the level of protection against too few parts desired. As shown in Section 2.2 the bias and variability in turn depend mostly on the coefficient of variation of the individual part weights, and the sample size used in the calibration step.

As an example, suppose $p = 25$ and we aim for 2% over the actual target. Figure 6 explores the effect of the coefficient of variation, and changing either the aimed for percent over target, or the calibration sample size on the chance the dribble counting will yield less than *n* parts. We focus on the problem of too few parts since in our application this is the major concern. The probability of too few parts are determined using (2), (3) and a normal approximation. Figure 6 is based on the assumption that $n = 500$ parts are desired, thought the results are not very dependent on the value for *n*. Notice the strong dependency on the coefficient of variation of the individual part weights.

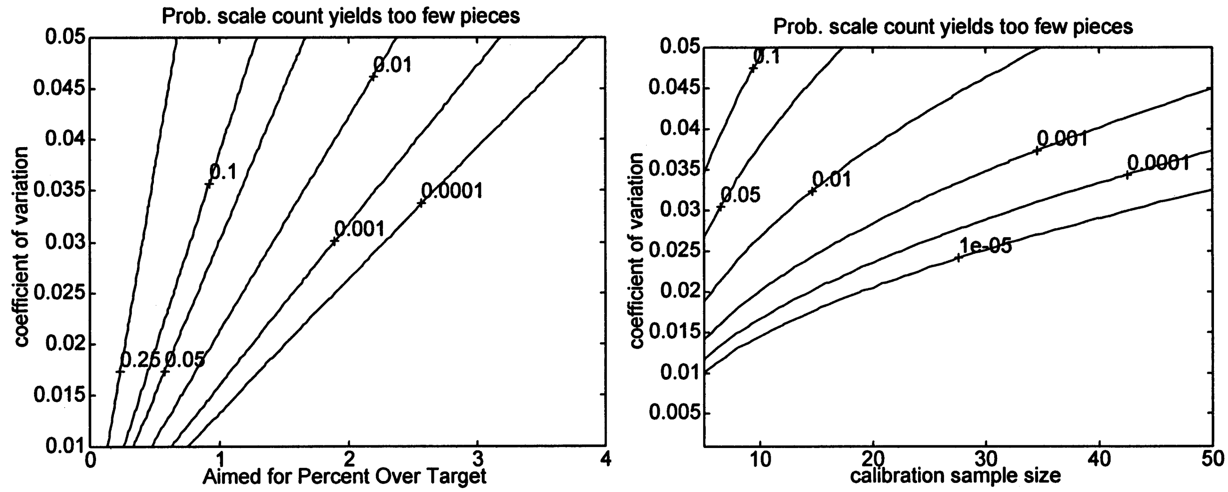

Figure 6: Contours of the estimated probability that scale counting yields too few parts
p=25 on left, aim for 2% over target on right

4.4 Multiple Scale Counting

As shown in Section 3.1, the rounding off to an integer number of parts usually increases the variance in the scale count by approximately $1/12$. However, when the variance (and bias) in the continuous count is small, say less than around 0.2, the discretization can reduce the variability (and bias) in the integer count dramatically. In most scale counting applications this is not important since the variability in the continuous count is relatively large. However, this large potential reduction in variability suggests an alternative strategy using multiple scale counts. Multiple scale counting is useful in situations where obtaining an accurate count is essential, yet the number of parts needed is too large to reliably count by hand, and automatic counting is not feasible. The idea is simple. To scale count n parts, we divide the n parts into k subgroups of around m parts each, where $km = n$. Then, rather than scale count all n parts at once we scale count each subgroup of parts separately and combine the results. By choosing m sufficiently small we can obtain count results that are extremely accurate. Of course there is a tradeoff between accuracy and effort required since with multiple scale counting k separate scale counts are needed and their results need to be combined.

To see that the use of multiple scaling counting must be useful in some circumstances consider two extremes: scale count all items together in one group (standard scale counting), and scale count each item individually. The latter approach is, of course, equivalent to individually counting the items which we wanted to avoid due to expense. It is clear that unless there is extremely large variability in the individual part weights that using the latter method will always result in the correct estimate for the total number of items, i.e. the method will have no estimation bias or variability. The former approach on the other hand, as we saw in Section 2, could have substantial bias and variability. Is there some compromise between these two extremes where the benefits of reducing estimation variability out weighs the disadvantage of the additional time and effort?

When using multiple scale counts, the estimate of the overall number of items would be given by

$$\hat{n}_{mult} = \sum_{j=1}^{k} round\left(\hat{t}_{m(j)} / \hat{\mu}\right) \qquad (6)$$

where $\hat{t}_{m(j)}$ is the observed weight of all the m items in the jth subgroup, $\hat{\mu}$ is the estimated average part weight obtained from the calibration step, and $mk=n$. Note that $\hat{\mu}$ is the same for all subgroups since the calibration step is performed only once. When all items are weighed together, as in standard scale counting, the estimate for the total number of items is given by (1). The estimators corresponding to (6) and (1) denoted $\sum_{j=1}^{k} round\left(\tilde{t}_{m(j)} / \tilde{\mu}\right)$ and $round\left(\tilde{t}_n / \tilde{\mu}\right)$ can be very different due to the rounding. Usually the latter is a better estimator since it has less variability. But in some circumstances, i.e. when m is small, the former has much less variability and bias than the latter estimator.

Obtaining the distribution of the estimator $\sum_{j=1}^{k} round\left(\tilde{t}_{m(j)} / \tilde{\mu}\right)$ is difficult since we need to consider the distribution of the estimator $\tilde{t}_m / \tilde{\mu}$ for each individual scale count and how the individual results could be combined. Some combinatorial algorithms to determine the distribution when n is relatively small have been developed by the authors.

As an example, consider a situation where we wish to scale count approximately 2000 items, the coefficient of variation of the individual part weights is approximately .01, and we use a calibration sample of size p=25. Figure 7 plots the standard deviation of the final count as a function of the number of subgroups we employ. We see that the standard deviation in the total count is quite large when we perform a single scale count of all the items. We obtain a substantial reduction in uncertainty in the final count by using as few as 20 scale counts of around 100 items each.

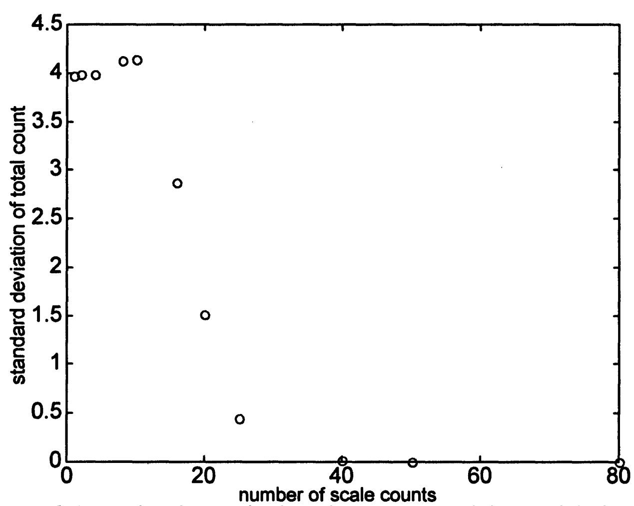

Figure 7: Expected Standard Deviation in Count with Multiple Scale Counting
n=2000, σ/μ=.01, p=25

Although the distribution of $\sum_{j=1}^{k} round\left(\tilde{t}_{m(j)} / \tilde{\mu}\right)$ is very complex, we can derive a simple rule of thumb to determine when multiple scale counting could be beneficial. The rule of thumb is derived from Figure 4 and an approximation for the variance of the scale count given by either expression (3), (4), (5) or some combination of these expressions. In Figure 4 we see that the discretization reduces the variability in the count whenever $Var\left(\tilde{t}/\tilde{\mu}\right)<0.2^2=0.04$. Combining this result with the approximation given by (3) suggests that multiple scale counting with subgroups of size at most $\sqrt{\frac{p}{\gamma^2 25}+\frac{p^2}{4}}-\frac{p}{2}$ will reduce the variability in the total count. For example, say p=25 and σ/μ=.01 the rule of thumb suggests that to be effective the subgroup size should be less than around 88 units. This matches closely the results given by Figure 7. Figure 8 plots contours of the maximum subgroup size suggested by the above rule of thumb. Figure 8 can be used to determine if multiple scale counting can be economically justified in any given example application.

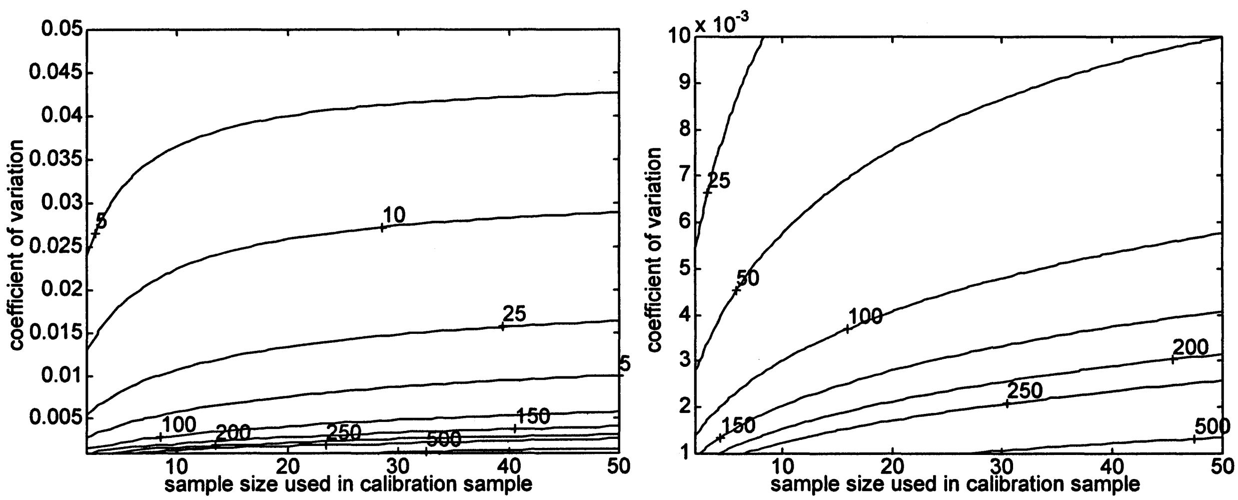

Figure 8: Contours showing the Maximum Subgroup Size for Effective Multiple Scale Counting

One potential problem with multiple scale counting is that the incorrect number of subgroups will be used. To avoid this problem we recommend finishing the multiple scale count with a scale count of all subgroups together. If the result of all subgroups together is close to the desired value this verifies that the correct number of subgroups was used.

5. Discussion and Recommendations

The results derived in this article suggest that scale counting can be an effective method of counting small parts. However, the accuracy of the count depends critically on a number of factors, including the sample size used to estimate the average part weight, the number of part we wish to count, and the coefficient of variation of the individual part weights. The effects of these factors are approximated in Section 2. The effect of poor measurement resolution and measurement variability can have a substantial negative effect on scale counting. However, the results show that for reasonably good measurement devices the negative effect of measurement variability and resolution is small. Measurement bias, on the other hand, could have a substantial effect, but can be avoided using differencing. Finally, the discretization used to obtain integer estimates of the count usually has little effect, but can be very beneficial in eliminating the bias and variability in the total count estimate in certain circumstances.

One important aspect of the results is that the standard deviation of the weights of the individual parts plays an important role in determining whether scale counting will be successful. However, estimation of this standard deviation is not possible with the current standard practice since in the calibration step p parts are weighed together. To estimate the standard deviation of the part weights individual parts or groups of parts must be weighed separately. This suggests that separate studies are needed in each application to determine how best to conduct the scale counting.

Acknowledgments

This research was supported, in part, by the Natural Sciences and Engineering Research Council of Canada. The authors would like to thank Maria Cruz and Jeffrey Chevers at TDS Automotive for allowing and encouraging studies of scale counting at TDS.

References

Anonymous (1981) "Industrial Weighing – Part 1: Electronic Counting Scales," *Modern Materials Handling*, 36, 50-59.

Cabuk S. and Springer M. (1990) "Distribution of the Quotient of Noncentral Normal Random Variables," Communications in Statistics: Theory and Methods, 19, 1157-1168.

Kotz S. and Johnson N.L. editors (1982) Encyclopedia of Statistical Sciences, John Wiley and Sons, New York.

Wolynetz M.S. (1979) "Maximum Likelihood Estimation from Confined and Censored Normal Data," *Applied Statistics,* 28, 185-195.